Can silicon change photonics?

Bahram Jalali*

University of California, Los Angeles, Department of Electrical Engineering, Los Angeles, CA 90095-1594, USA

Received 2 July 2007, accepted 29 November 2007
Published online 6 February 2008

PACS 42.55.Px, 78.66.Db, 78.66.Fd, 81.05.Cy, 81.05.Ea, 85.60.–q

*e-mail jalali@ucla.edu

The electronic chip industry embodies the height of technological sophistication and economics of scale. The industry mass produces complex circuitry, boasting over one billion components at such low cost that they appear in consumer products. Fabricating inexpensive photonic components by leveraging this mighty manufacturing infrastructure has been the impetus behind the development of silicon photonics. If it can be done economically and in an energy-efficient manner, empowering silicon with optical functionality will bring optical communications to the realm of computers where limitations of metallic interconnects are threatening the industry’s future. Guided by such visions and propelled by pioneering research conducted in the 1980s and 1990s, silicon photonics has enjoyed spectacular progress in the last six years. The critical size of photonic devices has been scaled to the 300 nm regime making more efficient use of wafer real estate. Optical amplification and lasing, once considered forbidden in silicon, have recently been demonstrated. High-speed and efficient electrical-optical conversion is being performed by production-worthy devices. Silicon’s nonlinear optical properties, enhanced by tight optical confinement in Si/SiO2 structures, are producing wavelength switching and generation, which are central functions in multiwavelength communications and signal processing. This article reviews these and other exciting progress being made in silicon photonics and identifies the challenges that remain on the path to commercialization.

1 Introduction The concept of silicon as a platform for integrated optics has its genesis in the pioneering works of Soref and Petermann, performed in the late 1980s and early 1990s. Key early findings that paved the path for much of the subsequent progress were (1) micrometer-size waveguides with single transversal modes can be realized despite the large refractive index difference between silicon and SiO2 [1], and (2) electro-optic switching can be achieved through modulation of free-carrier density [2]. These findings stimulated the activities in the 1990s that demonstrated many of the basic components needed in optical communication [3–7]. An early review of the field is described in Ref. [8].

Today’s intense activities in the field can be traced back to the technology boom of the late 1990s. Supported by significant private capital funding and driven by the need for low-cost components for the internet, silicon photonics emerged from this tumultuous era as a formal and respected discipline within the field of photonics. While the level of private funding diminished in the early 2000s, it was replaced with investments by large corporations and government agencies, the combination of which has fueled the spectacular progress witnessed in the last five years [9, 10].

Creating low-cost photonics, by exploiting the mighty chip manufacturing infrastructure, has been motivation for silicon photonics researchers. A fortuitous development was the emergence of silicon-on-insulator (SOI) as the platform of choice for high-performance CMOS. SOI also offers an ideal platform for creating planar optical circuits. The strong optical confinement offered by the high index contrast between silicon (n = 3.45) and SiO2 (n = 1.45) makes it possible to scale down the size of photonic circuits. Such lateral and vertical dimensions are required for economic compatibility with IC processing. In addition, the high optical intensity arising from the large index contrast (between Si and SiO2) makes it possible to observe nonlinear optical interactions, such as Raman and Kerr effects, in chip-scale devices. This realization has led to the development of optical amplifiers, optically pumped lasers, and wavelength converters – functions that until recently were considered to be beyond the reach of silicon.
The case for silicon photonics is broader than traditionally envisioned. The compatibility with CMOS notwithstanding, silicon has excellent material properties that are important in photonic devices. These include high thermal conductivity (∼10× higher than GaAs), high optical damage threshold (∼100× higher than GaAs), and high third-order optical nonlinearities (∼100× higher than optical fiber). Silicon is highly transparent from 1.1 μm to nearly 7 μm [11]. Furthermore, the absence of two-photon absorption at wavelengths larger than 2.25 μm renders silicon an excellent nonlinear optical material in the mid-wave IR spectrum. At the same time, entirely new functionality can be realized when electronics and photonics are combined onto the same chip. An example is the concept of the time-stretched analog-to-digital conversion [12]. This technique combines an optical time stretcher with an electronic digitizer and permits real-time capture of electrical waveforms whose bandwidth exceeds that of electronic analog-to-digital converters. This technique is described in more detail later in this paper.

The large volume of published papers and the length limitations of this article do not permit a comprehensive historical review of the field. Instead, this paper aims to describe the potential impact of this technology, recent breakthroughs that are keeping the technology in the limelight, and the challenges that remain before silicon photonics impact is fully realized.

2 Applications

The highest impact of silicon photonics may be in optical interconnection between digital electronic chips [13–17]. This technology addresses the communication bottleneck in VLSI electronics. An example is the Cell™ processor that is at the heart of Sony’s PlayStation 3 game console. Boasting nine processor cores, the chip has an internal computation power of 256 gigafloating point operations per second (GFLOPS) and communicates with the peripheral graphics processor and memory at data rates of more than 25 GBps [18, 19]. Such data rates challenge copper-based interconnects and represent a case where silicon chips with optical input/outputs will be highly desirable. A key finding is that for photonic interconnects to be advantageous over their copper counterparts, wavelength-division multiplexing (WDM) must be employed as shown in Fig. 1 [15]. The reason is that the minimum width of optical waveguides is limited by the optical wavelength [15]. Plasmonic waveguides can in theory overcome this limit but the losses for such waveguides are too high to be practical.

Figure 1 (online colour at: www.pss-a.com) Comparison of metal and optical interconnects for onchip communication. The figure of merit is in GBps/μm ps and represents the bandwidth normalized to wire width and latency time. Results show that optical interconnects will be advantageous only if wavelength-division multiplexing (WDM) is employed. The reason is that the minimum width of optical waveguides is limited by the optical wavelength [15]. Plasmonic waveguides can in theory overcome this limit but the losses for such waveguides are too high to be practical.

Bahram Jalali is a Professor of Electrical Engineering at UCLA. He is a Fellow of IEEE and the Optical Society of America (OSA), and recipient of R.W. Wood Prize from OSA. In 2005 he was elected into the Scientific American Top 50. From 2001–2004, he served as a consultant at Intel Corporation. He received the Bridegate 20 Award in 2001 for his entrepreneurial accomplishments and serves on the Board of Trustees of the California Science Center.
equalization and digital error correction, metal interconnects can address higher data rates, albeit at the cost of higher power dissipation (an excellent case study is the 10 Gigabit Ethernet over twisted pair copper also known as the IEEE 802.3ab standard). Therefore, for optical interconnects to replace their metal counterparts, they must be competitive in power. Power dissipation of computer and communication chips is a central issue facing the industry. Figure 2 shows that the power consumption per square foot in data centers is increasing at an exponential rate. Today, the power consumed by these centers represents nearly 2% of the electricity consumption in the U.S. This trend is creating formidable economic and logistic challenges for internet as well as utility companies.

Other applications for silicon photonics are being considered. For example, silicon photonics may be able to produce disposable mass-produced biosensors [20, 21]. One likely application is the so-called lab-on-a-chip in which both reaction and analysis are performed on a single device. Such sensors, along with integrated intelligence and wireless communication circuitry, may form nodes of an intelligent sensor network or environmental monitoring.

A near-term application is low-cost components for optical communication networks, such as the ones operating based on the SONET or the 10 Gigabit Ethernet standards. Low-cost photonics can bring power of optical networking to personal computers. With a computer’s copper networking cable replaced by an optical fiber, high-definition video can be effortlessly transferred through local area networks and between a computer and its peripherals.

An example of the start of the art is a prototype of an optical transceiver chip demonstrated by IBM that offers 160 Gbps of bandwidth (16 channels at 10 Gbps each) [22]. It consists of driver and receiver circuitry fabricated in a CMOS process. The chip is bump bonded in a flip-chip assembly to attach the separate laser and photodiodes. The chipset boasts low power dissipation of 2.5 W. Luxtera Corporation is also a leader in this field and has announced a 40 Gbps (4 channels at 10 Gbps each) which also integrated the photodetector onto the CMOS process [23].

Figure 2 Power consumption per square foot of area in internet data centers. The exponential increase of electrical power consumption represents an economical and logistic challenge for internet and utility companies.

A silicon photonic device that is in commercial production is Kotura Corporation’s 8-channel variable optical attenuator (VOA). The product is used for automatic channel equalization in Nortel Corporation’s products aimed at metropolitan networks [24]. The fiber-pigtalled product is fabricated on a SOI substrate using a dedicated fabrication facility.

3 Light modulation

The centrosymmetric arrangement of the silicon lattice does not lend itself to Pockels’ effect; a 2nd-order nonlinear effect that enables electro-optical modulation in LiNbO$_3$ and III–V semiconductors. This leaves the free-carrier dispersion effect as the simplest means to achieve fast modulation [25, 26]. Electrically modulated using a diode that straddles a waveguide (Fig. 3) the density of free carriers in the silicon waveguide changes both the real and imaginary parts of the refractive index leading to both electroabsorption and electrorefraction effects. This approach has been extensively studied and is described in excellent review articles about this topic [27].

An important milestone for modulator development is the 10 Gbit/s operation – a standard data rate for current long-haul (SONET OC192) and local-area networks (10 GbE). Researchers from Intel Corporation were the first to demonstrate silicon modulators operating at 10 Gbit/s [28]. The device bears resemblance to an MOS capacitor and consists of n-type crystalline silicon with an upper ‘rib’ of p-type silicon created by epitaxial lateral overgrowth (ELO). The n-type and p-type regions are separated by a thin insulating oxide layer. Upon application of a positive voltage to the p-type silicon, charge carriers accumulate at the oxide interface, changing the refractive-index distribution in the device. This index change, in turn, induces a phase shift that is converted to intensity modulation in a Mach–Zehnder interferometer. The advantage of this approach is that the MOS structure lends itself to inte-

Figure 3 (online colour at: www.pss-a.com) Canonical silicon photonic device consists of a silicon rib waveguide that resides on a buried oxide layer. The rib is straddled with a p–n junction diode. The diode controls the carrier (electron and hole) density in the waveguide and hence modulates its absorption. The most common use of this device is for electrical modulation of light. This structure also appears in nonlinear optical devices. Here, the reverse-biased diode reduces the nonlinear loss by removing the carriers that are unintentionally generated through two-photon absorption.
4 Photodetectors Silicon photodetectors operating at wavelengths below 1000 nm, where band-to-band absorption occurs, have been commercially available for several decades. They are also used, in conjunction with scintillators, as X-ray detectors that appear in computed tomography (CT) medical imaging equipment and airport luggage scanners.

The challenge as it relates to optical communications is that silicon is transparent in the fiber optic communication bands around 1300 nm and 1550 nm because the photon energies are less than the bandgap. With a smaller bandgap (0.7 eV vs. 1.1 eV for Si), germanium has strong absorption at these wavelengths and has been the main focus of researchers.

The lattice constant of Ge is 4% larger than that of Si and this poses a problem when Ge detectors are grown on a silicon substrate. The strain resulting from lattice mismatch builds up with film thickness and causes dislocation defects that increase the leakage current of the p–i–n photodetector. A second challenge is that Ge does not form a stable oxide, and the lack of a high-quality passivation layer for Ge makes it difficult to achieve a low dark current. Despite these challenges, excellent progress has been made towards Ge on silicon detectors [36] that are approaching III–V detectors in terms of speed and responsivity [37–40].

At high intensities, pure silicon can detect below-bandgap radiation via two-photon absorption (Fig. 4). Such detectors have been used as autocorrelators [41, 42] and as phase detectors in optical phase-locked loops that perform clock recovery and signal regeneration [43]. A TPA detector can be used as an optical power monitor and, in a closed-loop system, can form the basis of optical channel equalization in wavelength-division multiplexing (WDM) networks [44]. TPA-induced free-carrier absorption can also be used as an optical pulse compressor [45]. Here, the delayed build up of free carriers carves the optical pulse absorbing its trailing edge (Fig. 5). It has been shown that this phenomenon occurs inside a silicon electro-optical modulator allowing the device to be used as an intracavity modulator in an actively mode-locked laser [45].
5 Optical amplifiers and lasers

Silicon has an indirect band structure that means that the upper and the lower electronic states (conduction and valence bands) do not occur at the same value of crystal momentum. On the other hand, compared to the electrons, infrared photons used for optical communications have negligible momentum. Hence, for momentum to be conserved, the recombination of an excited electron needs to be mediated by emitting or absorbing a phonon. Such second-order radiative recombination events do not occur frequently and accordingly, radiative processes are insignificant compared to nonradiative recombination. Nonradiative recombination can be minimized by using high-purity samples, but even then, electrical to optical conversion efficiency is limited to only $10^{-3}$–$10^{-4}$ [45].

While the indirect bandgap is typically blamed for silicon’s inability to amplify light and hence to lase, the actual problem is more complex. It is best described by Fig. 6 that shows the calculated optical gain in Si and GaAs as a function of injected carrier density [46, 47]. Carriers injected to achieve population inversion also introduce free-carrier absorption (FCA). In GaAs, the high rate of direct transitions easily overwhelms the FCA loss resulting in significant net optical gain. In contrast, silicon’s low rate of radiative recombination is insufficient to overcome the FCA. The second problem in silicon is the high carrier density at which gain appears. This quantity, being proportional to the effective density of states, is reflected in the high band-edge effective mass of silicon. This is problematic, because at high carrier densities approaching $10^{19}$ cm$^{-3}$ Auger recombination (a nonradiative process) becomes significant with a lifetime that decreases as $N^{-2}$. Interestingly, if Ge was also plotted in Fig. 6, one would still observe negative gain but it would be much closer to the breakeven point than silicon.

Approaches aimed at overcoming or circumventing silicon’s optical amplification limitation belong to three main categories: (1) growth of III–V semiconductors on silicon, (2) using spatial confinement of the electron; (3) introducing rare-earth impurities as optically active dopants; and (4) using Raman scattering. Below, we briefly review the salient features of each approach and refer the reader to Ref. [48] for a more comprehensive review.

5.1 Epitaxial growth of III–V on silicon

A near-term approach for realizing a silicon-based laser is the heterogeneous integration of III–V on silicon. Direct growth of GaAs on silicon was extensively studied in the 1980s and has recently attracted renewed interest [49]. One noteworthy recent development is the room-temperature demonstration of an InGaAs quantum dot laser grown directly onto a silicon substrate by Bhattacharya’s group at the University of Michigan [50]. Another approach is the flip-chip bonding of a III–V gain element onto silicon demonstrated by Bower’s group at UCSB in collaboration with Intel [51, 52]. High temperatures encountered in CMOS processing and the fact that III–V elements behave as dopants in silicon pose challenges that must be overcome. However, if such process and material compatibilities can be established, these approaches stand an excellent chance of empowering silicon with electrically injected laser functionality.

5.2 Quantum confinement and erbium doping

GaP is a semiconductor that is used for light-emitting devices despite its indirect bandgap. The device works because the momentum-conservation requirement is relaxed when an electron is localized at a Nitrogen impurity site. According to the Heisenberg uncertainty principle, localization introduces ambiguity in momentum, a welcomed consequence that relaxes the momentum-conservation requirement.

This approach offers a path towards efficient light emission and amplification in silicon. Indeed, researchers at Brown University have reported evidence of lasing at cryogenic temperatures that is attributed to localization of electrons at defect sites that reside on silicon surfaces [53].
The sample under study consisted of two-dimensional array of nanometer-size holes etched into a thin film of silicon that resides on an oxide layer. It was cleaved (forming mirrors) and pumped optically. The optical input-output characteristics show a threshold behavior for emission at a wavelength of 1278 nm. Indicative of lasing, this behavior occurs at temperatures below 70 K. No such behavior was observed at the secondary emission wavelength of 1135 nm, which originates from the substrate. The emission spectrum measured at 10 K shows narrow peaks when the threshold pump intensity is increased supporting the claim that lasing has been achieved.

Another approach to implement quantum confinement in silicon has been the use of silicon nanocrystals that occur naturally in a silicon-rich oxide (SRO) thin film [54–56]. Such material can be pumped optically, or electrically using a tunnel diode, and has a spontaneous emission spectrum that lies in the range 800–900 nm.

The erbium-doped optical fiber is an established technology for light emission and amplification at the communication band of 1550 nm. Unfortunately, silicon is not a good host for Er, resulting in poor emission at room temperature. The reason is believed to be the back-transfer of energy from the excited Er ions to silicon and also the low concentration of Er that can be accommodated by silicon. Hence, Er doping of SRO, rather than Si, has been pursued as a mean to realize emission at the communication band of 1550 nm.

While optically pumped gain has been reported in SRO by several groups, the observations are highly dependent on how the sample was prepared with some showing gain and others exhibiting pump-induced loss [57, 58]. Consequently, questions remain regarding the nature of the experimental observations and the reproducibility of the results [57, 58]. This approach and issues surrounding it are discussed in excellent review articles by Fauchet [57] and Pavesi [58].

A simpler approach, and one that mimics Er-doped waveguide amplifiers (EDWA), is Er-doped glass waveguide, which can be deposited on a silicon substrate. In this case, the material is similar to SRO, except it contains no nanocrystals. To compensate for the low gain per unit length of Er-doped glass, high-Q-factor microdisk resonators have been used leading to very low threshold in optically pumped lasers [59].

### 5.3 Stimulated Raman scattering

An approach that can provide optical amplification and lasing in a manner that is independent of the bandstructure of the material is Raman scattering. The use of this approach in silicon was proposed in 2002 [60, 61] and soon led to the demonstration of the first silicon laser in 2004 [62]. The pulsed laser demonstration was followed by demonstration of continuous wave (CW) lasing in 2005 [63].

The first-order Raman scattering employed in these devices involves scattering of light from the zone-center optical phonons. In silicon, the zone-center phonon is triply degenerate with a frequency of 15.6 THz and bandwidth of approximately 100 GHz. This imposes a maximum information bandwidth of approximately 100 GHz that can be amplified. The Raman linewidth can be broadened through the use of multiple pump wavelengths or a single pump that has a linewidth larger than 100 GHz.

The main challenge in silicon Raman devices and other nonlinear optical devices is the loss caused by the free carriers that are generated by the high-intensity pump through TPA [64, 65]. Carrier lifetime determines the steady-state density of the generated carriers and hence the extent of the pump-induced loss. To obtain gain under steady-state pumping, a low lifetime or active removal of carriers must be employed. While bulk silicon has a catastrophically long lifetime of $10^{-3}–10^{-6}$ s, in thin SOI waveguides the lifetime can be as low as a few nanoseconds due to recombination at the Si/SiO$_2$ interface [66]. It is even lower in the case of submicrometer waveguides because of the recombination at the etched waveguide facets [67, 68]. Indeed, a modest amount of CW gain has been reported in deep submicrometer waveguides by Osgood’s group at Columbia University [69] where the impact of surface and interface recombination plays a critical role. The lifetime can be further reduced by introducing midgap states through high-energy irradiation and gold or platinum doping. The carrier density can also be reduced by using a reverse bias p–n junction to sweep the carriers out [64, 65] and CW gain has been achieved [69, 70]. Free-carrier screening of the junction electric field, a phenomenon reminiscent of high power saturation in photodetectors, limits the maximum pump intensity that can be used [71]. Furthermore, the diode can result in significant electrical power being dissipated on the chip.

By compensating for coupling and propagation losses, a Raman amplifier can make an impact in optical interconnects and low-cost photonic modules. However, the device length will have to be drastically lower than the centimeter-long device demonstrated so far. At an intensity of 100 MW/cm$^2$ (100 mW pump coupled into a 0.1 µm$^2$ waveguide) and with silicon’s Raman gain coefficient of ∼20 cm/GW, the gain will be ∼1 dB/mm. This low gain per unit length creates a challenge for miniaturization of Raman amplifiers, a prerequisite imposed by the economics of wafer manufacturing.

The chip size can be reduced by using meander-line waveguide geometry. Additionally, waveguide length will not be an issue if the device is used as a standalone discrete amplifier (similar to the role played by the Er-doped waveguide amplifier (EDWA)). In general, Raman gain increases linearly with pump intensity, whereas the TPA-induced free-carrier loss increases quadratically. Consequently, the gain peaks at a pump intensity above which free-carrier loss becomes significant. To obtain maximum total gain, the pump intensity along the amplifier length must be kept at an optimum value. By optimizing the pump-Stokes mode overlap using a cladding pump waveguide design, it has been shown that silicon Raman
amplifiers can offer extremely high gain [72, 73]. Since these amplifiers can provide gain outside erbium’s and III–V semiconductor’s gain spectra, silicon Raman amplifiers may represent an important class of waveguide amplifiers that complement EDWAs and semiconductor optical amplifiers (SOAs).

As in all lasers, the threshold pump power in silicon Raman lasers can be lowered by reducing the cavity propagation losses. The high index contrast of silicon waveguides results in larger scattering losses as compared with glass waveguides or other low-index waveguides. Using advanced processing and innovative coupler designs, the Intel team has recently demonstrated a ring cavity silicon Raman laser with a relatively low threshold pump power of 20 mW under CW operation [74].

5.4 Silicon image amplifier Recently, a new photonic device has been proposed that combines Raman amplification with the Talbot self-imaging phenomenon [9, 74–77]. The device consists of collinearly propagating pump and Stokes beams. The Stokes beam is amplified through energy gained from the pump beam, but at the same time, it is focused at periodic intervals along the waveguide including at the waveguide output (Fig. 7). This device has several applications, perhaps the most important of which is as an image preamplifier for laser remote sensing and imaging systems (Fig. 8). Here, a pixilated image is simultaneously amplified and focused as it propagates through the device. By elevating the incoming signal above the thermal noise level of the optoelectronic image sensor, the preamp will improve the sensitivity of laser-based imaging systems. The technology is best suited for operation in midwave infrared (MWIR) due to the lack of TPA-induced free-carrier effects. A unique feature of the physics of Raman interactions in multimode waveguides is that the Raman amplification process is accompanied by Raman-induced four-wave-mixing between various spatial modes of pump and image (Stokes) beams. This complex phenomenon places a limit on the maximum achievable gain, beyond which the image begins to distort. The image may also become distorted due to preferential amplification of Stokes modes that have the highest overlap with the pump. Such effects introduce a tradeoff between the gain and image quality, however, this tradeoff can be eliminated by selectively launching the pump beam into a single higher order waveguide mode [77].

6 Propagation and coupling losses, and optical filters The high index contrast between silicon and SiO$_2$ or air makes it possible to reduce the optical mode size to approximately 0.1 $\mu$m$^2$, i.e. on the same scale as typical dimensions encountered in CMOS VLSI. However, the high index also results in large propagation losses when light scatters from the surface roughness that occurs during the etching of silicon. Typical losses range from 0.2–3.0 dB/cm with higher losses occurring in smaller waveguides. While postprocessing, such as thermal oxidation, can be used to reduce the roughness on the waveguide sidewall and to minimize the losses [78–80], the tradeoff between waveguide size and loss generally exists.

Surface morphology has been shown to also improve by high-temperature annealing in a hydrogen ambient [81]. Recognizing that some of the roughness in etched structures originates from the nonsmooth edges of the photore sist pattern that is used as the etch mask, resist reflow was employed to improve the surface morphology to the extent that losses were dominated by surface-state absorption [82]. It is therefore expected that, similar to the case in electronic devices, surface passivation will play an increasingly important role in photonics.

Another challenge is the problem of coupling light between a single-mode fiber that has an effective area, $A_{\text{eff}} \sim 50$ $\mu$m$^2$, and a scaled silicon waveguide with $A_{\text{eff}} \sim 0.1$ $\mu$m$^2$. Fortunately, efficient coupling between the fiber and the so-called “silicon wires” can be achieved using the inverse-taper approach (Fig. 9) [78, 83–85]. This technique relies on the gradual expansion of a core-guided
mode into a much larger cladding guided mode and coupling loss as low as 0.2 dB has been demonstrated [83]. Another approach is surface gratings etched onto the silicon [23, 86, 87]. A curved grating geometry simultaneously performs phase matching and focusing [87] and coupling losses of 1 dB are obtained in experiments. The attractive feature of the grating approach is the normal-incidence geometry. Being compatible with conventional circuit-testing equipment, it makes it perform rapid onchip validation without the need for cleaving and polishing of chips facets.

In the context of low-cost components for optical networks, optical filters used for wavelength multiplexing (WDM) and demultiplexing represent an important class of passive devices. Arrayed waveguide grating, the workhorse of WDM communication, was demonstrated in SOI as early as 1997 [7]. More recently, microring and microdisk structures have been extensively explored [78, 88] including devices with MEMS tuneability [89]. The challenge in such structures is achieving a high extinction ratio, a task that is difficult due to fabrication-induced errors. Recently, an add-drop filter with an impressive 50 dB extinction ratio has been demonstrated by Ippen’s and Smith’s groups at MIT [90]. The SiN waveguide device achieved this performance without postfabrication trimming and offers flat-top bandpass performance with 2 dB of drop loss.

7 Other devices In this section, we briefly review several other recent and significant developments. A saturable absorber, the basic building block of mode-locked lasers, has been demonstrated in silicon [91]. Realized as a silicon/SiO2/germanium Bragg reflector, the device has been used to demonstrate mode locking of an Er-Yb: glass laser. When used with a silicon gain medium, it can enable an all-silicon mode-locked laser.

Wavelength conversion, a central function in all-optical routing of data packets, has been demonstrated using both the Kerr [92, 93] and Raman nonlinearities [94, 95]. Kerr-based conversion is most suitable for channel conversion while Raman provides a means to convert data between the widely spaced 1300 nm and the 1550 nm wavelength band. Using the Kerr effect under phase-matched conditions, parametric gain has been demonstrated [96]. The Kerr nonlinearity has also been used to demonstrate continuum generation, a technique that makes it possible to generate multiple wavelength channels on the chip from a single wavelength off-chip source [97–99]. Such capability bodes well for WDM based optical interconnects.

Photonic bandstructure devices represent an important subfield within silicon photonics. The high index contrast of Si/SiO2, and Si/air renders SOI the ideal platform for such devices. Photonic crystals have tremendous potential, as evident by recent demonstrations of unique functionalities, such as cavities with ultrasmall mode volumes [100], the superprism effect [101, 102], self-guiding or the so-called supercollimation [103], ultracompact waveguide bends [104], and slow light [105]. The main challenge with such devices is the large optical loss encountered in deeply etched periodic structures. Fundamentally, this arises from surface scattering that is made more pronounced due to the high surface-to-volume ratio in these structures. Another challenge is the sensitivity to fabrication-induced errors in device geometry. Nevertheless, photonic bandstructure devices may find a role in future systems, not as a replacement for more conventional devices, but rather by providing unique functionalities that are not offered by them.

8 Outlook Silicon photonics is experiencing an astonishing rate of progress, and it is expected that technological progress will continue in the foreseeable future. The question of whether the promises of this technology will come to fruition will depend on how well it can function within the constraints of the chip industry. The constraints are two-fold, and comprise of economic issues and the thermal-dissipation problem.

The first economic requirement is that products must have a high-volume market. If so, the large nonrecurring engineering (NRE) cost of the CMOS process is amortized among the large number of devices sold, resulting in a low per-unit cost. Being CMOS-based does not translate into low cost, unless the product has a high-volume market. Such economics pose a challenge for companies that use a state-of-the-art CMOS process to develop products for a nascent market.

8.1 Three-dimensional integration If photonics is to merge with VLSI electronics, it cannot significantly increase the chip area, otherwise it will violate the economics of silicon manufacturing. To be sure, the more efficient use of wafer real estate has been fueling the relentless scaling of CMOS to smaller feature lengths (Moore’s law) because it directly impacts the profit margin of chip manufac-
MOS transistors on top of buried photonic circuits \[106, 107\].

Vertically coupled photonic devices as well as 3D integration through semitransparent masks and has been used to realize multicore processors \[108\]. This major shift in the semiconductor roadmap was fueled by the rapid rise in chip heat dissipation \[109\] with numbers approaching 100 W/cm². Among photonic components, lasers (and laser driver circuits) are the most power-hungry devices. The lack of an electrically pumped silicon laser, to date, dictates an architecture where the light source remains off-chip. By placing the “optical power supply” off-chip, this architecture is in fact preferred as it removes a main source of heat dissipation. Furthermore, the performance degradation of injection lasers at high temperatures will be a major obstacle to their integration onto the “hot” VLSI substrate, even when such silicon lasers are demonstrated. Nevertheless, creating an electrically pumped silicon laser or a practical LED is a novel and worthy goal as such devices can impact optoelectronic displays and other high-volume products.

Modulators, amplifiers, photodetectors, and perhaps wavelength converters are destined to be integrated on-chip. Among these devices, the modulator and the amplifier have the highest power dissipation. Similar to bipolar transistors, carrier-injection-type optical modulators suffer from static power dissipation. On the other hand, depletion-mode free-carrier modulators, as well as new devices based on the quantum-confined Stark effect in Si/Ge quantum wells \[110, 111\], will have low static power dissipation.

8.2 Power dissipation

Recently, the microprocessor industry opted to halt increasing the clock speed in favor of multicore processors \[108\]. This major shift in the semiconductor roadmap was fueled by the rapid rise in chip heat dissipation \[109\] with numbers approaching 100 W/cm². Among photonic components, lasers (and laser driver circuits) are the most power-hungry devices. The lack of an electrically pumped silicon laser, to date, dictates an architecture where the light source remains off-chip. By placing the “optical power supply” off-chip, this architecture is in fact preferred as it removes a main source of heat dissipation. Furthermore, the performance degradation of injection lasers at high temperatures will be a major obstacle to their integration onto the “hot” VLSI substrate, even when such silicon lasers are demonstrated. Nevertheless, creating an electrically pumped silicon laser or a practical LED is a novel and worthy goal as such devices can impact optoelectronic displays and other high-volume products.

Figure 10 (online colour at: www.pss-a.com) Ultimate solution to efficient use of wafer real estate is 3D integration of electronics on top of buried photonics. The example shown here is an MOS transistor fabricated on top of a subsurface microdisk resonator. The fully monolithic process employs multilayers of SOI, formed by patterned oxygen implantation, in which photonic devices are first formed in the buried silicon layer \[106, 107\].

The technology employs a multilayer SOI structure in which the photonic devices are first formed in the buried silicon layer \[106, 107\]. The process makes use of oxygen implantation, in which photonic devices are first formed in the buried silicon layer. The buried layer is patterned by implantation of oxygen through masks residing on the surface layer (Fig. 11). The surface silicon layer is then used for conventional CMOS processing.

The use of silicon wire waveguides with sub 1000 nm transverse dimensions partly addresses this requirement. This approach notwithstanding, the ultimate answer will be 3D integration of electronics and photonics, a preliminary example that has recently been demonstrated (Fig. 10) \[106, 107\]. The technology employs a multilayer SOI structure in which the photonic devices are first formed in the buried silicon layer. The buried layer is patterned by implantation of oxygen through masks residing on the surface layer (Fig. 11). The surface silicon layer is then used for conventional CMOS processing.

Figure 11 (online colour at: www.pss-a.com) 3D sculpting of microphotonic structures consisting of high-index Si and low-index SiO₂ sections. The process makes use of oxygen implantation through semitransparent masks and has been used to realize vertically coupled photonic devices as well as 3D integration of MOS transistors on top of buried photonic circuits \[106, 107\].

8.3 Harvesting the energy consumed by two-photon absorption

TPA-induced free-carrier loss is the main problem in nonlinear optical devices, including Raman- and Kerr-type devices. A reverse-biased diode is an effective method to eliminate this nonlinear loss by removing free carriers. Unfortunately, this solution can come at the expense of significant heat dissipation. As a case study, about one Watt of electrical power had to be dissipated to achieve ∼4 dB of CW optical gain and to produce ∼8 mW of output from the Raman laser \[62\]. In the context of CMOS-based photonics, this should be compared to about one μW of power that is dissipated by a MOS transistor in a VLSI circuit. Energy-efficient methods for dealing with the TPA-induced loss are hence a priority. Recently, it was shown that it is possible to eliminate the nonlinear loss with negative electrical power dissipation \[112, 113\]. In this approach, the TPA-generated carriers are swept out by the built-in field of the junction diode, yet by doubling as a two-photon solar cell, the device delivers electrical power by harvesting the photon energy that was consumed by the TPA process. This effect is visible in the typical measurements shown in Fig. 12. Here, the current–voltage curve under optical pumping shows a reverse current flowing even under forward diode voltage, as long as the voltage is below the diode turn-on threshold. The reverse current is due to the sweep out of the TPA-generated free carriers, yet because the product of the voltage and current is negative, the device has negative power consumption.

Free-carrier density can also be reduced by employing lifetime-reduction techniques that are normally used in producing silicon step-recovery diodes and fast photoconduc-
The author’s own belief is that silicon photonics is a technology whose time has come. It stands to impact a number of industries ranging from computing and communication to biomedicine. Fueled by recent government and private-sector investments, the technological progress has been nothing short of spectacular. Going forward, the technology’s faith will be governed not by technological breakthroughs alone, but also by careful attention to the economics of chip manufacturing and the power-dissipation issues that lie on the path to commercial success.

Acknowledgement Much of the progress described in this article would not have been possible without the generous support from the Defense Advanced Research Project Agency (DARPA). We are particularly indebted to Dr. Jag Shah of DARPA for his visionary support of silicon photonics.

References